Human Molecular Genetics, 2003, Vol. 12, No. 21 **2733–2743**
DOI: 10.1093/hmg/ddg314

Association of extreme blood lipid profile phenotypic variation with 11 reverse cholesterol transport genes and 10 non-genetic cardiovascular disease risk factors

Alfredo Morabia[1,*], Eftihia Cayanis[2], Michael C. Costanza[1], Barbara M. Ross[2], Maria Sol Flaherty[2], Gabriela B. Alvin[2], Kamna Das[2] and T. Conrad Gilliam[2]

[1]Division of Clinical Epidemiology, Geneva University Hospitals, 25 rue Micheli-du-Crest, 1211 Geneva 14, Switzerland and [2]Columbia Genome Center, and Departments of Genetics and Development, and Psychiatry, Columbia University, 1150 St Nicholas Avenue, New York, NY 10032, USA

Received April 30, 2003; Revised August 12, 2003; Accepted September 3, 2003

This study explored the genetic basis of the combination of extreme blood levels of HDL-C and LDL-C, a well-studied endophenotype for CVD, which has several attractive features as a target for genetic analysis: (1) the trait is moderately heritable; (2) non-genetic risk factors account for a significant but still limited portion of the phenotypic variance; (3) it is known to be moderated by a number of gene products. We exhaustively surveyed 11 candidate genes for allelic variation in a random population-based sample characterized for known CVD risk factors and blood lipid profiles. With the goal of generating specific etiological hypotheses, we compared two groups of subjects with extreme lipid phenotypes, from the same source population, using a case–control design. Cases ($n = 186$) were subjects, within the total sample of 1708 people, who scored in the upper tertile of LDL-C *and* the lowest tertile of HDL-C, while controls ($n = 185$) scored in the lowest tertile of LDL-C *and* the upper tertile of HDL-C. We used logistic regression and a four-tiered, systematic model building strategy with internal cross-validation and bootstrapping to investigate the relationships between the trait and 275 genetic variants in the presence of 10 non-genetic risk factors. Our results implicate a subset of nine genetic variants, spanning seven candidate genes, together with five environmental risk factors, in the etiology of extreme lipoprotein phenotypes. We propose a model involving these 14 genetic and non-genetic risk factors for evaluation in future independent studies.

INTRODUCTION

Cardiovascular disease (CVD) risk is influenced by several well-established risk factors, such as body mass index (BMI), an indicator of overweight and obesity, blood lipids, diabetes, and blood pressure (1). These are intermediate phenotypes, correlated among themselves (1) and having their own genetic and environmental determinants, including diet (2,3), nutrition (4), hormones (5,6), smoking (7), alcohol intake (8), and physical activity (9,10).

Studies in humans and mice indicate that both the type and quantity of blood-borne lipids are predictive of cardiovascular health or disease and that a relatively large number of proteins is involved directly or indirectly in the transport, maintenance and elimination of blood lipids, including high and low density lipoprotein cholesterol (HDL-C and LDL-C, respectively) (11–15). We identified 11 of these proteins (depicted in Fig. 1) with a well-known collective activity in the reverse transport of cholesterol from peripheral cells to the liver. Genes coding for these proteins that have been implicated with blood lipids comprise: ATP-binding cassette protein 1 (ABCA1) (16); apolipoprotein A1 (APOA1) (17); apolipoprotein E (APOE) (18,19); cholesterol-ester transfer protein (CETP) (20–22); endothelial lipase (EL) (23); hepatic lipase (HL) (24); lecithin–cholesterol acetyl transferase (LCAT) (25); low density lipoprotein receptor (LDLR) (26); lipoprotein lipase (LPL) (27,28); phospholipid transfer protein (PLTP) (29); and scavenger receptor class B type I (SR-BI) (30). With the possible exceptions of APOE and LPL, the vast majority of studies attempting to correlate candidate gene variants with

*To whom correspondence should be addressed. Tel: +1 41223729552; Fax: +1 41223729565; Email: alfredo.morabia@hcuge.ch

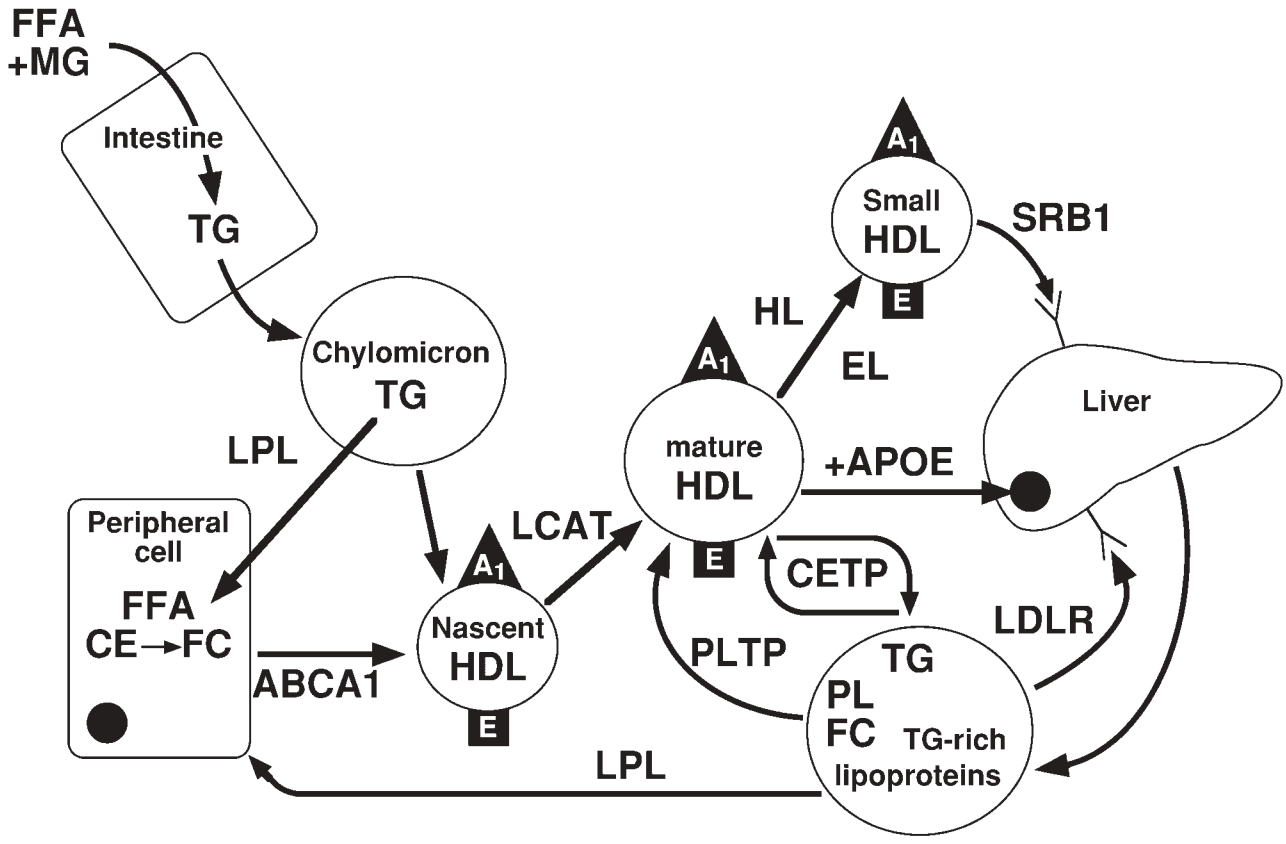

Figure 1. Graphical representation of the reverse cholesterol transport (RCT) metabolic pathway, from which we selected 11 genes: ABCA1, APOA1, APOE, CETP, EL, HL, LCAT, LDLR, LPL, PLTP and SR-BI. Dietary free fatty acids and monoglycerides form triglycerides (TG) in the intestine, which are transported by chylomicrons. TG are also transported as very low-density lipoproteins (VLDL) formed in the liver. LPL releases free fatty acids from chylomicrons and VLDL in peripheral tissues (heart, muscle, adipose). The TG-depleted 'chylomicron remnants' take on APOE and serve, in the macrophages, as a basis for the construction of APOA1 and APOE-containing HDL particles. These nascent HDL particles interact with peripheral cells and acquire cholesterol and phospholipids through a transport process facilitated by ABCA1. Nascent HDL evolves into mature HDL in part via the PLTP-mediated transfer of phospholipids and free cholesterol from TG-rich lipoproteins to HDL, and via the esterification of free cholesterol within the HDL particle by the LCAT enzyme. These cholesteryl esters (CE) form the core of the mature HDL, which can be further enriched with APOE prior to their uptake as particles in the liver. CEs can also be selectively transferred, in exchange for TG, to TG-rich lipoproteins through the action of CETP. These TG-rich lipoproteins can then undergo hepatic endocytosis via the action of LDL-receptors. HL and EL hydrolyze HDL-TG and phospholipids, thereby reducing the size of HDL and stimulating the SR-BI-mediated selective hepatic uptake of CE.

common, heritable traits have yielded either false positive findings, or else real but weak genotype–phenotype associations. This shortcoming is inherent in association studies based on limited numbers of candidate gene variants (31–34). In an attempt to overcome this limitation, Knoblauch *et al.* (35) typed 29 SNPs in six genes (APOE, CETP, HL, LCAT, LDLR, LPL) among 732 subjects from 222 nuclear families and analyzed haplotypic effects. The associations between each gene and blood lipoproteins were more significant for haplotypes than for single nucleotide polymorphisms (SNPs).

In the present study we explore the associations between the 11 target genes shown in Figure 1 and blood HDL-C and LDL-C concentrations. We made no attempt to model the reverse cholesterol transport (RCT) pathway, *per se.* However, in contrast to Knoblauch *et al.* (35), we combined the variants across the 11 genes in our statistical models. Moreover, since statistical power to detect lipid-related genetic variation may be reduced when the confounding effects of known non-genetic risk factors are not taken into account, we advantageously measured genotypic variation in a sample that had been characterized in detail for established CVD risk factors.

We considered using as phenotypes HDL-C, LDL-C or triglycerides (TG) individually, or some combination of all three. We opted finally for a combination of HDL-C and LDL-C because some of the protein products of the 11 target genes are implicated in the regulation of both HDL-C and LDL-C levels. It could be argued that selection upon either HDL-C or LDL-C alone would have provided a less complex, more genetically homogeneous phenotype, but the choice of one or the other measure might have captured only part of the combined phenotypic effects of the targeted genes. We systematically surveyed exons for major allelic variants in a subsample of 95 subjects. We then compared the allelic and genotypic distributions of the ensuing collection of polymorphic variants (mostly SNPs) in two groups of subjects with extreme lipid phenotypes, from the same source population, using a case–control study design. Cases were those 186 individuals who, within a total sample of 1708 people, scored in the upper tertile

of LDL-C *and* the lowest tertile of HDL-C distributions, while controls ($n = 185$) scored in the lowest tertile of LDL-C *and* the upper tertile of HDL-C distributions. The case–control subjects also tended to score in the corresponding extreme tertiles of TG, because HDL-C, LDL-C and TG levels are highly correlated in the extreme tails of their distributions. All the analyses were simultaneously adjusted for known non-genetic cardiovascular risk factors, and the remaining residual variance was modeled using an analytic strategy starting with a total of 275 genetic variants spanning the 11 candidate genes.

RESULTS

For ease of reference, Table 1 provides a flowchart of the steps involved in obtaining the results described in this section and indicates (see last column) the tables and figures where detailed data are shown either in the paper or in its Appendix on the authors' web site (www.epidemiology.ch, choose 'publications').

Resequencing: identification and characterization of SNPs

A total of 275 SNPs were identified (Table 2 and Appendix Table A). Of these, 190 (69%) mapped to introns, flanking regions, or 5′ and 3′ flanking/untranslated regions (UTR), while 85 (31%) mapped to exons. Among the exonic SNPs, 34 (40%) coded for non-synonymous amino acids. There were 139 (51%) 'common' SNPs with rare allele frequency $\geq$3%. Nucleotide diversity ($\times 10^{-4}$) was 7.2 overall, but 10.5 in promotor, 6.5 in exons, 7.7 in introns, 7.4 in 3′-UTR and 7.7 in 5′-UTR sequences (36).

Comparison of resequencing and remaining case–control sample rare allele frequencies

Excluding the nine LDLR SNPs retrieved from a public database plus nine other resequenced common SNPs that could not be assayed in the remaining cases and controls for technical reasons, 121 of the 139 common SNPs could be compared between the resequencing subsample and the remaining case–control sample. The rare allele frequencies of these 121 SNPs were very similar in these two groups (frequency difference between the $n = 95$ resequencing sample and the $n = 276$ remaining case–control sample: median = 0%, Wilcoxon signed-rank test $P = 0.26$; histogram in Appendix Figure A).

Non-genetic characteristics and lipid profiles of the case–control sample

Detailed distributions of the categorical and continuous nongenetic covariates among the 186 cases and 185 controls have been presented recently elsewhere (37). As expected due to extreme phenotypic selection, the cases were heavier, less educated, more sedentary, and, on average, 4 years older (all $P < 0.03$). These factors were included as confounders in the analyses. Cases had also higher total cholesterol (mmol/l: cases, 6.61; controls, 5.01, $P < 0.0001$) and higher TG (geometric mean: cases, 1.55; controls, 0.85, $P < 0.0001$).

Comparison of case–control sample and excluded subjects

A comparison of the final case–control sample ($n = 371$) versus the 79 extreme phenotypic subjects who were excluded due to insufficient blood samples remaining for DNA analyses is shown in Table 3. Apart from being more sedentary (76 versus 63%, $P < 0.03$), the excluded subjects were very similar to the final case–control sample on all the other non-genetic characteristics and lipid profile measurements investigated. It is unlikely that this observed difference in sedentarity prevalence would modify the results substantially because sedentarity was not a major confounder in the analyses.

Single-SNP association study

In our first analysis step (tier 1 modeling) we identified individual SNPs which were distributed differently in cases and controls, after adjustment for the nongenetic factors. First in tier 1a, of the 130 valid common SNPs, 51 had individual, adjusted case–control odds ratios (ORs) either $\geq$1.50 ('atherogenic SNPs'), or $\leq 1/1.50 = 0.67$ ('atheroprotective SNPs'), independently of statistical significance, or had an SNP $\times$ gender interaction (nominal), $P \leq 0.10$, *and* fulfilled at least one of the first two conditions for men or women or both (most often differently in each gender) ('gender-dependent SNPs', overriding the other two groups). Then in tier 1b we excluded one SNP from each pair of highly correlated SNPs, and were also forced to eliminate SNPs with rare allele frequencies <5% because their regression coefficients were unstable in the logistic models. At the end of tier 1, 37 SNPs were eligible for multi-SNP analyses (10 atherogenic + 17 atheroprotective + 10 gender-dependent SNPs). These 37 common SNPs comprised 13 exonic [nine coding (six non-synonymous + three synonymous)] and 24 non-exonic SNPs (Appendix Table B).

Multiple SNP association study

In our second analysis step (tier 2 modeling) we eliminated statistically redundant SNPs (*only*) using stepdown selection *within* each of the three tier 1 SNP groups from logistic models (the non-genetic covariates were forced to remain in the three models). There were six atherogenic + eight atheroprotective + six gender-dependent SNPs retained after tier 2 (Table 4).

In our third analysis step (tier 3 modeling) we *pooled* the 20 retained SNPs from tier 2 across all three SNP groups and performed a further stepdown selection of SNPs *only* (non-genetic covariates again forced to remain in the logistic three models). There were one atherogenic + six atheroprotective + three gender-dependent SNPs retained after tier 3 (Appendix Table D).

In our fourth analysis step (tier 4 modeling) the 10 SNPs retained after tier 3 became eligible for the final model, but they competed with each other *and* with the 10 non-genetic covariates in the final stepdown selection. The tier 4 final model (see below) comprised nine SNPs (+two SNP $\times$ gender interactions) and five non-genetic factors (BMI, gender, alcohol, current smoking, age). In the cross-validation, there was considerable overlap between the five 80% random subsample SNPs and the

Table 1. Steps of gene selection, population sampling, and statistical modeling strategy for reverse cholesterol transport (RCT) SNP data (numbered tables/figures/references are in the text; lettered tables/figures are in the Appendix)

Step	Content	Number of remaining SNPs	Number of eliminated SNPs	See
RCT genes	Define biological model based on 11 genes	—	—	Figure 1
Population-based random sample	Men and women residents of Geneva, Switzerland, ages 35–74 years, 1999–2000 ($n = 1708$ untreated for hypercholesterolemia)	—	—	References (9,44)
Blood lipid extreme phenotype case–control sample	Combined HDL-C *and* LDL-C tertile groups: 186 cases—low HDL-C + high LDL-C; 185 controls—high HDL-C + low LDL-C [$n = 371$ ($= 450 - 79$ excluded)].	—	—	Table 3
	Statistical power (186 cases + 185 controls)	—	—	Figure B
SNP detection	Resequencing of 5′, 3′ flanking/UTR regions and all exons in subsample of cases + controls ($n = 95$).	256	—	Table A
	Public database SNPs for LDLR	19	—	Reference (47)
		Total = 275	—	Tables 2 + A
Genotyping	All common SNPs ($\geq$3%) assayed in remaining cases and controls ($n = 276$).	139	136	Table A
	Valid genotyping.	130	9	
	Comparison of allele frequencies for resequenced ($n = 95$) versus genotyped data ($n = 276$)	121 (all genes but LDLR)	—	Figure A
Tier 1[a] modeling *before* Tier 1b exclusions	Retain SNPs with adjusted single-SNP case–control OR $\geq$1.50 or OR $\leq$0.67	(see below) Total 51	79	Table B
Tier 1a[a] SNP groups	Atherogenic (AG) (OR $\geq$1.50); atheroprotective (AP) (OR $\leq$0.67); gender-dependent (GD) (see text)	AG: 14	—	
		AP: 23	—	
		GD: 14	—	Table B
Tier 1b[a] exclusions (multicollinearity)	Eliminate one SNP when $\lvert r \rvert \geq 0.7$, within or between the AG, AP, GD SNP groups	AG: 12	2	Table B
		AP: 18	5	
		GD: 13	1	
Tier 1b[a] exclusions (rare SNPs)	Eliminate rare SNPs (rare allele $\leq$5%) within the AG, AP, GD SNP groups	AG: 10	2	Table B
		AP: 17	1	
		GD: 10	3	
Tier 1[a] modeling	Total SNPs *after* tier 1b exclusions	Total 37	Total 14	Table B
Tier 2[a] modeling	*Within* AG, AP, GD groups (nominal) $P = 0.10$ stepdown selection of SNPs (10 non-genetic covariates forced into models)	AG: 6	4	Tables 4 + C
		AP: 8	9	
		GD: 6	4	
		Total 20	Total 17	
Tier 3[a] modeling	Pool tier 2 SNPs across AG, AP, GD groups. Stepdown SNP elimination (nominal $P = 0.05$) (10 non-genetic covariates forced into model)	AG: 1	5	Tables 4 + D
		AP: 6	2	
		GD: 3	3	
		Total 10	Total 10	
Tier 4[a] modeling	Stepdown elimination (nominal $P = 0.05$) of 10 Tier 3 SNPs *and* 10 non-genetic covariates.	AG: 1	0	Table E
		AP: 5	1	Table 5
		GD: 3	0	Table 6
		Covariates: 5	Covariates: 5	
	Bootstrap final model.	9 SNPs + 5 covariates	1 SNP + 5 covariates	
	Cross-validate case-control predictions[b]	9 SNPs + 5 covariates	1 SNP + 5 covariates	
Tiers 1–4[a] modeling	Cross-validate *entire* modeling process[a]	—	—	Tables B, C, D

AG, atherogenic SNPs, AP, atheroprotective SNPs; GD, gender-dependent SNPs.

[a]Tiers 1–4 models internally cross-validated in five random subsamples each comprising 80% of the subjects.

[b]Tier 4 model case–control predictions internally cross-validated in five random subsamples each comprising 20% of the subjects excluded from modeling. Appendix tables and figures are on the authors' web site: www.epidemiology.ch, choose 'publications'.

Table 2. Eleven reverse cholesterol transport (RCT) genes and numbers of SNPs assayed

Genes					SNPs assayed				
Acronym	Name	GenBank no.	Chromosome position	Numbers of exons	Total[a]	Exonic	Coding	Intronic	Common[b]
ABCA1	ATP-binding cassette	600046	9q22–q31	50	88	19	4	69	53
APOA1	Apolipoprotein A–I	107680	11q23	4	4	1	1	3	2
APOE	Apolipoprotein E	107741	19q13.2	4	7	3	2	4	5
CETP	Cholesteryl ester transfer protein	118470	16q21	16	27	8	5	19	14
EL	Endothelial lipase	603684	18q12.1–q12.3	10	31	4	3	27	15
HL	Hepatic lipase	151670	15q21–q23	9	37	17	7	20	15
LCAT	Lecithin–cholesterol acetyl transferase	245900	16q22.1	6	4	2	0	2	2
LDLR	Low-density lipoprotein receptor	143890	19p13.2	18	19	12	3	7	9
LPL	Lipoprotein lipase	238600	8p22	10	24	9	3	15	13
PLTP	Phospholipid transfer protein	172425	20q12–q13.1	12	17	4	3	13	6
SR-BI	Scavenger receptor class B type I	601040	Chr. 12	13	17	6	3	11	5
	Totals			152	275	85	34	190	139

[a]Except for LDLR (see text), all SNPs were identified by resequencing a subsample 95 subjects (see text).
[b]With nine exclusions for technical reasons, 130/139 common SNPs (rare allele frequency ≥3%) were assayed in 371 subjects with extreme lipid phenotypes.

total sample SNPs within each of the four modeling tiers (Appendix Tables C, D and E).

Final multi-SNP logistic model

The final tier 4 bootstrap-estimated ORs and confidence intervals are presented in detail in Table 5. The nine retained SNPs spanned seven genes. Each SNP was associated with case–control status over and beyond the associations of the nongenetic factors. The two gender-dependent SNPs exhibited qualitatively different associations by gender: HL 3b−279 was protective in men (OR = 0.34, $P < 0.04$), but not in women (OR = 2.86, $P = 0.13$), while SR-BI A350A was protective in men (OR = 0.28, $P < 0.04$), but borderline deleterious in women (OR = 4.98, $P = 0.053$).

That our adjustment procedures increased the statistical power to detect SNP effects in this study was demonstrated by refitting and bootstrapping the final model with the nine SNPs, but *without* the non-genetic covariates other then gender. Except for the two bootstrapped gender-dependent SNP unadjusted ORs for women (HL 3b−279: 1.34; SR-BI A350A: 2.25), all of the other unadjusted SNP ORs were biased towards unity (*APOE2*: 0.24; PLTP 1b+26: 0.60; ABCA1 50b.3038: 0.41; LPL S447X: 0.46; ABCA1 32b+30: 2.38; LDLR2: 0.61; HL 1b−280: 0.54; HL3b−279 for men: 0.63; SRBI A350A for men: 0.36). The latter unadjusted estimates were as much as two times smaller than their adjusted counterparts, although all of them still had bootstrapped $P < 0.05$. On the other hand, four of the unadjusted SNPs (including the HL and SR-BI exceptions for women) had bootstrapped $P > 0.05$ (PLTP 1b+26: $P = 0.09$; LDLR2: $P = 0.13$, HL 3b−279 for women: $P = 0.62$; SRBI A350A for women: $P = 0.12$).

Also of note, the cross-validation subsample models tended to retain ABCA1 1b−1126 (+gender interaction) instead of the HL and SR-BI SNPs (+gender interactions). It was not surprising, given the statistical power of the study, that the largest modeling discrepancies occurred for gender-dependent SNPs. Augmenting the tier 4 model with ABCA1 1b−1126 (+ gender interaction) resulted in much-attenuated associations for the two HL SNPs. Thus, although Table 5 shows the most parsimonious, best model derived from the total case–control sample, the cross-validation results suggested that the selection of ABCA1 1b−1126 versus the two HL SNPs may be sample-specific.

Overall, the final tier 4 logistic model accounted for 60% of the case–control variance, correctly classifying 79% of the cases and 78% of the controls. The respective mean classification rates for the five cross-validation subsamples were 72 and 76% (Table 6). As discussed in further detail below, the final model needs to be validated in a future independent sample. The non-genetic risk factors were associated with phenotype as expected based upon previous findings, which suggestively bodes well for the general applicability and putative significance of our findings once an independent replication of our model is obtained. However, we caution that development of an appropriate permutation test to correct for possible multiple testing artifacts, or better, replication in an independent sample, will be required to finally address questions of statistical significance.

Swiss-born subjects only

In order to assess potential population stratification effects on the findings, we refitted the final tier 4 model to the Swiss-born subjects only, who comprised over half the sample (76 cases and 89 controls with complete data). With the exception of the SNP × gender interaction terms for HL 3b−279 and SR-BI A350A, which had to be omitted because of small sample sizes, both the magnitudes and directions of all the SNP main effect OR associations among the Swiss-born subjects were entirely consistent with those based on all subjects (not shown in a table).

DISCUSSION

This study explored the genetic basis of the combination of extreme blood levels of HDL-C and LDL-C, a well-studied endophenotype for CVD, which has several attractive features as a target for genetic analysis. First, the trait is known to be moderately

Table 3. Non-genetic characteristics and lipid profiles of the case–control sample (CCS) compared with subjects excluded due to insufficient blood for DNA analyses

		CCS ($n = 371$)	Excluded ($n = 79$)	
Categorical variable		n[a] (%)	n[a] (%)	P[b]
Case–control status	Cases	186 (50)	40 (51)	
	Controls	185 (50)	39 (49)	0.94
Gender	Men	170 (46)	41 (52)	
	Women	201 (54)	38 (48)	0.33
Education	University	135 (36)	25 (32)	
	Secondary	210 (57)	47 (59)	
	Primary	26 (7)	7 (9)	0.67
Country of birth	Switzerland	200 (54)	40 (59)	
	Others[c]	117 (31)	23 (21)	
	Mediterranean[c]	54 (15)	16 (20)	0.45
Cigarette smoking	Never	174 (47)	32 (41)	
	Ex-smoker	104 (28)	21 (27)	
	Current	89 (24)	26 (33)	0.27
Physical activity	Active[d]	137 (37)	19 (24)	
	Sedentary[d]	234 (63)	60 (76)	0.029
Continuous variable		*Mean* (*SD*)	*Mean* (*SD*)	P[b]
Age (years)		50.6 (9.8)	49.2 (10.3)	0.18
BMI (kg/m^2)		24.8 (3.9)	25.0 (3.6)	0.47
Dietary fat (%)		35.1 (7.0)	35.2 (6.8)	0.82
Dietary fiber (g/day)		16.1[f] (0.5)	14.5[f] (0.5)	0.080
Alcohol drinking (g alcohol/day)[e]		10.6[f] (1.0)	12.8[f] (0.5)	0.17
Lipid profile (mmol/l)	Total cholesterol	5.81 (1.0)	5.85 (1.0)	0.75
	HDL-cholesterol	1.40 (0.4)	1.36 (0.4)	0.40
	LDL-cholesterol	3.81 (1.1)	3.88 (1.1)	0.59
	Triglycerides	1.15[f] (0.5)	1.13[f] (0.6)	0.68

[a]Totals may vary due to missing values.
[b]k-category variables: χ^2 test ($k-1$ d.f.) P-value; continuous variables: minimum P-value for Student's t and Wilcoxon rank sum tests.
[c]Others: mostly France (approximately one-third), remainder <5% each; Mediterranean: Italy, Portugal, Spain.
[d]Active (sedentary): ≥10% (<10%) of total energy in activities requiring ≥4 × basal metabolic rate.
[e]Total (beer + wine + hard liquor) among drinkers [% CCS/excluded: 88%/82% ($P = 0.21$)].
[f]Geometric mean (SD of log data).

heritable. In German twins, heritability was 0.61 for HDL-C and 0.59 for LDL-C (38). Second, non-genetic risk factors account for a significant but still limited portion of the phenotypic variance. And third, intense biochemical and physiological investigations have identified a number of genes whose protein products have impacts on HDL and LDL bound cholesterol.

This is perhaps the first time that some ordering of the relative contributions of both genetic and non-genetic variables to the total variance of a lipid trait has been demonstrated. Nine SNPs (across seven genes) and five non-genetic factors remained associated with the extreme phenotypic lipid profiles in the final, most parsimonious model. BMI was the single most important determinant of 'caseness', with an adjusted r^2 of 0.27. Next in importance were *APOE2* (+0.05 to cumulative r^2) and PLTP 1b.+26 (+0.03), followed by (mostly low to moderate) alcohol drinking (+0.07) and (current) cigarette smoking (+0.03). After ABCA1 50b.3038 (+0.03 to cumulative r^2) and LPL S447X (+0.02), age (+0.01) and gender (+0.02) were added. From a biological perspective, the selected nongenetic factors make sense. Obese patients typically have dyslipidemia (4). Approximately half of the protective effect of alcohol may be mediated through increased levels of HDL-C (8). Smoking is associated with lower HDL-C and higher LDL-C (7). There are well-established age and gender-related differences in lipid metabolism, endothelial function, and risk of atherosclerosis (5). Our results also reflect the careful selection of biologically relevant candidate genes, and the effective conditioning of their putative effects upon known risk factors. They are consistent with our previous findings using single candidate gene approaches in the full $n = 1708$ sample, either for the association of APOE with HDL-C (9), or for the non-association of the SNP in position −629 of the promotor region of CETP with either HDL-C or LDL-C (39).

The case–control design based on phenotypically extreme subjects is a critical element of the study and merits several comments. Cases represented just over 10% of our sample with the most atherogenic HDL-C and LDL-C profiles, while controls comprised just over 10% of the most atheroprotective HDL-C and LDL-C profiles. For a fixed sample size, this design has more statistical power to detect genetic effects for the selected trait than one based on the total sample (40). The reason is that marginal effects estimated from the comparison of subjects with extreme phenotypes are likely to incorporate epistasis or interactions between genes and covariates, if these exist, biasing effect sizes upward (41). The statistical power to detect individual SNP effects is therefore increased. On the other hand, the absence of subjects with intermediate phenotypes reduces the statistical power to test for epistasis that may contribute to the phenotypic variation. Given the modest sample size of our study we therefore did not attempt to study gene–gene or gene–covariate interactions, except for gender which we planned by design. Had it been feasible to conduct our case–control study with a much larger sample, the outset of the analysis would have been the most appropriate stage at which to assess epistasis and would have involved all 130 common SNPs. We are currently working on ways to address this problem in a non-selected sample, but to the best of our knowledge it remains unsolved and most studies focusing on even a limited number of SNPs have not adequately addressed the issues of gene–gene and gene–environment interactions.

The current study had several salient strengths. The population-based random sample was selected from 1708 residents of Geneva, Switzerland, each of whom was evaluated for known risk factors for CVD and atherogenic lipid profiles. Methods to measure several of these factors were developed and validated in the same target population (42,43). The protein products of each of our 11 candidate genes were known to affect blood cholesterol levels. We completely sequenced the coding portion of each candidate gene (a total of 152 exons) along with the immediately adjacent non-coding DNA segments corresponding to both the 5′ and 3′ regions of each gene in a subset of 95 individuals to enhance the identification of common gene variants. We analyzed in the case–control study an exhaustive set of candidate genetic variants (130 SNPs of a total of 275). SNP effects proved to be enhanced by the adjustment for nongenetic covariates (a phenomenon

Table 4. Reverse cholesterol transport (RCT) SNPs retained *within* groups of atherogenic, atheroprotective, and gender-dependent SNPs using (nominal) $P = 0.10$ stepdown logistic model SNP selection (10 non-genetic covariates forced into model)

	Entry step[b]	r^2 adjusted[c]	Correct classification (%)[d]			Final model significance[c]	
			Total	Cases	Controls	*P*	OR (95% CI)
Atherogenic							
SNPs in final model[a]							
None (covariates only)	0	0.41	73	74	73		
APOE4	1	0.44	76	74	78	0.0009	3.58 (1.69–7.56)
CETP 1b.−629	2	0.46	75	74	76	0.0085	2.50 (1.26–4.93)
APOA1 2b.+33	3	0.47	75	75	76	0.0246	2.40 (1.12–5.15)
ABCA1 32b.+30	4	0.49	74	74	74	0.0374	2.31 (1.05–5.07)
LPL 6b.+82	5	0.50	75	72	78	0.0450	2.08 (1.02–4.25)
EL 10b.337	6	0.51	77	74	80	0.0812	1.92 (0.92–3.98)
Atheroprotective							
SNPs in final model[f]							
None (covariates only)	0	0.41	74	72	75		
APOE2	1	0.49	76	77	75	<0.0001	0.10 (0.04–0.28)
PLTP 10b.+70	2	0.53	77	78	77	0.0005	0.13 (0.04–0.41)
LPL S447X	3	0.56	77	76	79	0.0006	0.22 (0.10–0.53)
ABCA1 50b.3038	4	0.58	77	77	77	0.0007	0.15 (0.05–0.45)
CETP 6b.−56	5	0.61	79	79	78	0.0055	0.34 (0.16–0.72)
HL 5b.98	6	0.62	78	78	79	0.0446	0.43 (0.19–0.98)
HL 1b.−280	7	0.63	78	78	77	0.0248	0.44 (0.22–0.90)
LDLR2 (rs2228671)	8	0.64	77	78	76	0.0621	0.45 (0.20–1.04)
Gender-dependent							
SNPs in final model[g]							
None (covariates only)	0	00.41	73	71	75		
PLTP 1b.+26 + gender interaction	1	0.43	74	74	74	0.6215^{m}	0.77^{m} (0.27–2.19)^{m}
	2	0.45	73	71	75	0.0009^{w}	0.11^{w} (0.03–0.40)^{w}
ABCA1 50b.2037 + gender interaction	3	0.45	74	73	75	0.0519^{m}	2.96^{m} (0.99–8.82)^{m}
	4	0.46	74	71	76	0.6592^{w}	0.80^{w} (0.31–2.12)^{w}
EL 10b.504 + gender interaction	5	0.46	74	73	75	0.0369^{m}	0.35^{m} (0.13–0.94)^{m}
	6	0.47	76	73	79	0.3614^{w}	1.50^{w} (0.63–3.58)^{w}
SR-BI A350A + gender interaction	7	0.47	75	73	78	0.0263^{m}	0.30^{m} (0.10–0.87)^{m}
	8	0.48	77	75	79	0.1363^{w}	2.24^{w} (0.78–6.48)^{w}
ABCA1 1b.−1126 + gender interaction	9	0.48	77	75	79	0.1296^{m}	2.19^{m} (0.79–6.06)^{m}
	10	0.50	75	74	77	0.0312^{w}	0.29^{w} (0.09–0.89)^{w}
HL 3b.−279 + gender interaction	11	0.50	75	73	77	0.1034^{m}	0.42^{m} (0.15–1.19)^{m}
	12	0.52	76	75	77	0.1099^{w}	2.22^{w} (0.84–5.91)^{w}

[a]Subjects with complete data in tier 2 analysis of atherogenic SNPs: 159/186 cases, 160/185 controls (total 319/371).
[b]Order of forced-forward entry of SNPs retained in model after (nominal) $P = 0.10$ backward elimination.
[c]Adjusted r^2 = raw r^2/max(r^2), where: Raw $r^2 = 1 - [L(0)/L(\text{fitted})]^{2/n}$; L = likelihood (null versus fitted model), max $(r^2) = 1 - L(0)^{2/n}$.
[d]One-step approximation to leaving-one-out method. Case classification cutpoint: adjusted logistic probability of being a case ≥0.50.
[e]p = single SNP *P*-value (1 d.f. χ^2) adjusted for covariates and other SNPs in final model; OR (95% CI) = final model covariate-adjusted case/control SNP odds ratio (95% confidence interval).
[f]Subjects with complete data in tier 2 analysis of atheroprotective SNPs: 145/186 cases, 145/185 controls (total 271/371).
[g]Subjects with complete data in tier 2 analysis of gender-dependent SNPs: 142/186 cases, 145/185 controls (total 287/371); for each gender-dependent SNP main effect listed, the corresponding SNP × gender interaction term was included in the model; both main effect and interaction terms were used to calculate gender-specific ORs.
[m,w]Results shown separately for men[m] and women[w] for gender-dependent SNPs.

termed 'negative confounding' in epidemiology). Finally, we employed bootstrapping and cross-validation techniques to address some of the biases expected from our analytical procedures. On the other hand, the study also had its limitations. It was exploratory in nature. Our logistic regression modeling strategy systematically attempted to minimize the set of SNPs over multiple models without correction. Thus, we accept that some parts of the findings are likely to be false positive. However, our study succeeded in generating hypotheses to guide further experimentation, either in an extension of our current sample, or in other independent samples.

In conclusion, we have reported on an exhaustive survey of genetic variation spanning 11 candidate genes, all of which are known to produce proteins that play key roles in the homeostatic regulation of blood lipid profiles related to cardiovascular health and disease. The study compared subjects with extreme atherogenic or atheroprotective lipid profiles from a major, ongoing epidemiological study of a population-based random sample in Geneva, Switzerland. Each individual was thoroughly characterized for known risk factors for CVD and related blood lipid profiles. Our findings implicate a subset of nine of the 275 genetic variants, spanning seven candidate genes, together with five of the 10 environmental factors, in the etiology of the complex trait in our sample. We have proposed a model involving these 14 risk factors that merits evaluation in future independent studies.

Table 5. Contributions of reverse cholesterol transport (RCT) SNPs and nongenetic covariates to case–control lipid profile status in tier 4 logistic model (nominal) $P = 0.05$ stepdown selection of both SNPs and covariates, with bootstrap validation of final model *P*-values and adjusted odds ratios

SNPs and covariates retained in final model	Entry step[a]	r^2 adjusted[b]	Correct classification (%)[c]			9-SNP + 5-non-genetic covariate tier 4 final logistic model[d]	
			Total	Cases	Controls	*P*	OR (95% CI)
BMI (kg/m^2)	1	0.27	71	69	73	<0.0001	1.65 (1.43–1.98)
APOE2	2	0.32	74	74	74	0.0002	0.14 (0.03–0.33)
PLTP 1b.+26	3	0.35	72	73	72	0.0053	0.27 (0.08–0.61)
Alcohol: medium + high	4	0.38	74	77	72	<0.0001	0.02 (0.001–0.08)
Alcohol: low	5	0.42	73	74	72	0.0011	0.11 (0.02–0.30)
Current smoker	6	0.45	76	78	74	0.013	6.42 (2.06–16.6)
ABCA1 50b.3038	7	0.48	78	80	76	0.0032	0.20 (0.05–0.48)
LPL S447X	8	0.50	77	77	78	0.031	0.41 (0.14–0.86)
Age (years)	9	0.51	76	77	76	0.072	1.04 (0.997–1.09)
Gender (women)	10	0.53	78	79	78	0.012	0.11 (0.05–0.50)
ABCA1 32b.+30	11	0.54	77	77	78	0.030	3.80 (1.20–9.83)
LDLR2 (rs2228671)	12	0.55	78	77	80	0.042	0.38 (0.11–0.89)
HL 1b.−280	13	0.56	77	77	78	0.039	0.46 (0.18–0.94)
HL 3b.−279	14	0.56	77	75	78	0.039[m]	0.34[m] (0.08–0.92)[m]
HL 3b.−279 × Gender	15	0.57	78	78	79	0.13[w]	2.86[w] (0.79–7.74)[w]
SR-BI A350A	16	0.57	78	78	78	0.034[m]	0.28[m] (0.05–0.82)[m]
SR-BI A350A × Gender	17	0.60	78	79	78	0.053[w]	4.98[w] (1.03–15.9)[w]

Analyses based on 149/186 cases, 152/185 controls (301/371) with complete data on all covariates and 10 tier 3 selected SNPs.
[a]Order of forced-forward entry of SNPs and nongenetic covariates retained in model after (nominal) $P = 0.05$ backward elimination (pre-bootstrap).
[b]Adjusted $r^2 =$ raw $r^2/\max(r^2)$, where: raw $r^2 = 1 - [L(0)/L(\text{fitted})]^{2/n}$, $L =$ likelihood (null versus fitted model), max $(r^2) = 1 - L(0)^{2/n}$ (pre-bootstrap).
[c]One-step approximation to leaving-one-out method. Case classification cutpoint: adjusted logistic probability of being a case ≥0.50 (pre-bootstrap).
[d]Bootstrapped single-SNP or covariate odds ratio (OR) with 95% CI (bootstrap percentile confidence interval) adjusted for all other SNPs and covariates in model ($B = 2000$ bootstrap replicates). Test-based *P*-value (assuming H_0: OR = 1) derived from bootstrapped log(OR).
[m,w]Results shown separately for men[m] and women[w] for gender-dependent SNPs.

Table 6. Tier 4 model estimated correct classification rates, with internal cross-validations in five 80:20% random subsamples

Subgroup	Total sample[a]	Correct classification rates (%)					Subsample mean
		Random subsample[b]					
		A	B	C	D	E	
Cases	78	80	68	68	59	83	72
Controls	79	76	72	69	75	87	76
Total	78	78	70	69	67	85	74

[a]One-step approximation to leaving-one-out method; case classification cutpoint: adjusted logistic case probability ≥0.50.
[b]Cross-validation of model based on 80% random subsample obtained by classifying the 20% random subsample of excluded subjects (same cutpoint as in footnote a).

MATERIALS AND METHODS

Subjects

The population, sample recruitment and data collection methods have been described in detail elsewhere (9,44). In brief, survey respondents were randomly selected within gender/age strata throughout January 1999 to December 2000 from the resident population of Geneva, Switzerland aged 35–74 years. The present study sample comprises 846 men and 862 women not under treatment for hypercholesterolemia (total $n = 1708$ untreated study subjects).

A subset of 450 phenotypically extreme subjects was selected from the untreated subjects on the basis of their HDL-C and LDL-C concentrations falling in the lowest and highest tertiles (T1, T3 = 33.3rd, 66.7th percentiles) separately by gender. The tertile boundaries (mmol/l) for HDL-C and LDL-C were (men/women): HDL-C, T1 = 1.06/1.33, T3 = 1.32/1.63; LDL-C, T1 = 3.54/3.22, T3 = 4.30/4.04 (to convert to mg/dl, multiply by 38.6).There were 226 (13.2%) subjects with HDL-C ≤ T1 and LDL-C > T3 (low HDL, high LDL), which was considered an atherogenic phenotype. At the other extreme, there were 224 (13.1%) subjects with HDL-C > T3 and LDL-C ≤ T1 (high HDL, low LDL). However, 79 subjects had to be excluded from further study because of too little or no blood sample remaining for genetic analyses. The initial blood samples from these excluded subjects had not been used for any other studies. The final sample size was $n = 371$, comprising 186 'cases' (atherogenic phenotypes) and 185 'controls' (non-atherogenic phenotypes), comprising 11.4% each of the 1629 subjects with available DNA data.

Survey procedures

Respondents completed questionnaires and were examined in a mobile epidemiology clinic by trained health technicians. Blood pressure, height and weight were measured and venous blood was drawn. Diet and physical activity were recorded using semi-quantitative food and physical activity frequency questionnaires previously developed and validated in the same population (42,43). Total plasma cholesterol and TG and their concentrations in the lipoprotein subfractions were assayed in fasting blood (Bayer Technicon Diagnostics, Brussels, Belgium). Total genomic DNA was extracted from EDTA blood (Gentra Puregene blood kit, BioConcept, Allschwil, Switzerland). LDL-C (mmol/l) was calculated as (total cholesterol − HDL-C − TG/2.2) (45).

Laboratory procedures

Primers for genomic amplification were chosen outside the exon sequence. Each exon, along with intronic 5′ and 3′ ends (total 67 948 bp), was amplified separately and sequenced. Amplification primers are available upon request.

The entire coding DNA and the immediately adjacent noncoding DNA, including both 5′- and 3′-UTR and a portion of each intron, were assayed first in a purposefully selected resequencing subsample of $n = 95$ subjects in whom the largest genetic variability was anticipated: 48 subjects had a non-atherogenic lipid profile and were sedentary (see below), hence were expected to have a 'protective' genetic constitution, while the 47 other subjects had an atherogenic lipid profile but were physically active, hence were expected to have a 'deleterious' genetic constitution.

DNA from the resequencing subsample was amplified across exons (excluding LDLR, see below) using the Perkin Elmer Gene Amp PCR system 9600 as described in detail elsewhere (46). PCR amplification products were purified and sequenced in both directions, using big dye-terminator chemistry and 373-XL automated sequencers (ABI). Sequence comparisons were determined using the Sequencher Program version 4.1.2 (Gene Codes) and verified by visual inspection. Final determinations were made based on the presence of two peaks, each of which was approximately half the size of the corresponding homozygous peak. Manual confirmation was required due to complications from sequence irregularity and preferential PCR amplification of one allele over the other. For LDLR, we retrieved the information from public databases, mostly from Cargill *et al.* (47).

Statistical power for SNP detection

The sequencing component of this study had 99.9% (95%) power for detecting a SNP with rare allele frequency $\geq$5% ($\geq$1%), based on 95 subjects (48). We therefore decided to assay all SNPs with rare allele frequency $\geq$3% in the remaining cases and controls. Variants identified were assayed in the larger sample using a modified template-directed dye-terminator incorporation with fluorescence polarization (TDI-FP) detection (Acycloprime-FP SNP Detection Kit of PerkinElmer Life Sciences, Inc.). Primers to amplify and detect SNPs are available upon request.

Adjustment of single- and multi-SNP odds ratios (OR) for covariates and SNP coding

We used logistic regression models for assessing single and simultaneous SNP effects adjusted for multiple nongenetic covariates. The covariates were: gender, age (years), body mass index (BMI, kg/m^2), education [reference group = primary (<9 years schooling), secondary, university ($\geq$13 years and Swiss baccalaureate)], country of birth {reference = Switzerland + all other [approximately one-third France, remainder (mostly Northern Europe) <5% each)], Mediterranean (Italy, Spain, Portugal)}, cigarette smoking {reference = [never (<100 lifetime cigarettes) + exsmoker (quit $\geq$1 year pre-interview)], current}, alcohol drinking (g alcohol/day) [reference = none, low (men/women: 1–40/1–20), medium + high (men/women: 41+/21+)], total dietary fat (%), dietary fiber (g/day), and daily energy expenditure {reference = active [$\geq$10% total energy expended in physical activities requiring four or more times the basal metabolic rate (4BMR) (49)], sedentary (<10% in 4BMR)}. Overall, the 10 non-genetic covariates were expressed with 12 variables.

Each SNP was coded 0 if the rare variant was absent, and 1 if present (as heterozygote/homozygote). This coding assumes the rare allele effect is dominant. As homozygotes for the rare allele were extremely rare, it is essentially the same thing as an additive model, consistent with most evidence about the genetics of many quantitative traits. The correlation between two SNPs was assessed by Pearson's r statistic, equivalent to a standardized linkage disequilibrium (LD) coefficient (36). We used r mainly to avoid multi-collinearity in the logistic modeling (see below), not to assess LD, *per se*.

The outcome of the logistic regression was the probability of being a 'case', that is, of having an atherogenic lipid profile. The odds ratio (OR) of being a case (atherogenic) versus control (non-atherogenic) with 95% confidence interval (CI) was calculated. All covariates were entered as linear terms either as continuous or dummy indicator variables as listed above. The adjustment for covariates was done to neutralize potential confounding effects of variables that are related to the outcome and are distributed differently in cases versus controls.

Single- and multi-SNP adjusted logistic models

The individual SNP data were analyzed first with crude (unadjusted) case/control ORs, and next with single-SNP ORs from logistic models adjusted for the non-genetic covariates in which the outcome was case–control status and each SNP was the explanatory variable of interest. We then employed a four-tiered modeling strategy, summarized in Table 1, to investigate the simultaneous associations of the SNPs with extreme lipid profiles:

Tier 1 modeling strategy

Tier 1a. The individual SNPs were divided into three mutually exclusive groups: SNPs with individual case–control adjusted odds ratios (OR) $\geq$ 1.50 were considered 'atherogenic', and those with OR $\leq 1/1.50 = 0.67$ were considered 'atheroprotective'. SNPs with gender interaction (nominal) $P \leq 0.10$ *and* for which at least one of the latter conditions was satisfied by either men or women or both (usually differently in each gender), were considered 'gender-dependent' (overriding the first two SNP groups), and the interaction term was used in subsequent modeling to obtain the corresponding gender-specific results.

Tier 1b. Further exclusions *within* the three Tier 1a SNP groups were made for: (i) highly correlated SNPs ($|r| \geq 0.7$, within or between groups); (ii) SNPs with rare allele frequency $\leq$5% (due to small sample sizes, although we could *detect* SNPs with frequencies $\geq$3% in the resequencing subsample with excellent power, the logistic regression model coefficients were unstable for SNPs with frequency $\leq$5%). In the presence of high correlation between an exonic and an intronic/flanking region/UTR SNP, we retained the exonic SNP for modeling whether it coded for a synonymous or non-synonymous amino

acid change (31). For highly correlated non-exonic SNPs, arbitrary choices of which SNP to include in the modeling process were made. There was one exception for EL: the coding exon 3b.53 (T111I) SNP was not initially classified as gender-dependent (interaction $P = 0.21$). However, because of very high correlation ($r = 0.96$) with the non-coding gender-dependent intronic 4b.+98 SNP (interaction $P < 0.05$), T111I was *substituted* for 4b.+98 in the modeling.

Tier 2 modeling strategy

Within each of the three tier 1a SNP groups, all tier 1-retained SNPs were eligible for stepdown selection using (nominal) $P = 0.10$ in a logistic model in which the 10 non-genetic covariates were forced to remain. For gender-dependent SNPs, the SNP × gender interaction term was eligible to be removed if the SNP main effect was in the model, and an SNP main effect removal implied simultaneous removal of the interaction term. Except for occasional missing data, the three tier 2 within-SNP group analyses were performed on the *same* sample of subjects.

Tier 3 modeling strategy

All tier 2-retained atherogenic, atheroprotective, and gender-dependent SNPs were *jointly* eligible for stepdown selection using (nominal) $P = 0.05$ in a logistic model in which the 10 non-genetic covariates had been forced to remain.

Tier 4 modeling strategy

All tier 3-retained SNPs *and* the 10 non-genetic covariates were *jointly* eligible for stepdown selection using (nominal) $P = 0.05$ in a logistic model.

The predictive contributions of the SNPs and non-genetic covariates retained in the tiers 2–4 logistic models were assessed by the cumulative adjusted coefficient of determination (squared multiple correlation), r^2 (50). For this purpose the stepdown-retained SNPs/covariates were forced into the model in the order that they would have been entered using forward selection. We also calculated the cumulative overall and case–control-specific correct classification rates using a one-step approximation to the leaving-one-out method (case-cutpoint: adjusted case probability >0.50).

In order to validate the final tier 4 logistic regression model SNP and non-genetic covariate *P*-values and ORs, a standard Monte Carlo bootstrap (resampling with replacement) procedure based on 2000 replicates was applied (51). The potential impact of population stratification on the findings was assessed by refitting the tier 4 logistic model to Swiss-born subjects only and comparing the results to those from the final model based on all subjects.

Internal cross-validation of tiers 1–4 logistic models

In order to assess replicability of the logistic modeling process, the *entire* four-tiered modeling strategy was internally cross-validated by applying it to five 80 : 20% random partitions of the dataset. The cases and controls were first randomly ordered separately within gender, and each of the four resulting randomly permuted subgroups was divided into fifths and recombined into five different 80 : 20% split datasets. The *complete* tiers 1–4 modeling strategies were then applied to each of the five 80% subsamples to provide comparisons with the total sample results. Finally, each 20% subsample excluded from the modeling process was classified into cases and controls using the tier 4 model based on the corresponding 80% subsample.

Statistical power of the case–control association study to detect single-SNP ORs $\neq 1$

We described above the statistical power to *detect* SNPs in the resequencing sample of $n = 95$ subjects. Here we describe the statistical power to detect non-null SNP *effects* with the case–control sample of $n = 371$ subjects. The latter is a function of the true effect size of the SNP and of its rare allele frequency (Appendix Figure B). For example, statistical power (1-β) was 80% (given two-tailed $\alpha = 5\%$) to detect single-SNP ORs ranging from 1.75 to 2.00 (or 0.50 to 0.57) for rare allele frequencies $\geq$17.5%, from 2.10 to 2.30 (or 0.44 to 0.48) for frequencies between 10 and 15%, and from 2.30 to 2.90 (or 0.35 to 0.44) for frequencies between 5 and 10%.

ACKNOWLEDGEMENTS

We thank Victoria Haghighi and Joe D. Terwilliger for comments on an earlier version of the manuscript. This project was supported by grants from the Swiss National Science Foundation (nos 3200-054097 and 3200-068275.02), and by funding from the Columbia Genome Center. This study was part of the Geneva programme for metabolic disorders.

REFERENCES

1. Anonymous (2001) Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA*, **285**, 2486–2497.
2. Ye, S.Q. and Kwiterovich, P.O. Jr (2000) Influence of genetic polymorphisms on responsiveness to dietary fat and cholesterol. *Am. J. Clin. Nutr.*, **72**, 1275S–1284S.
3. Ordovas, J.M. (2002) Gene–diet interaction and plasma lipid responses to dietary intervention. *Biochem. Soc. Trans.*, **30**, 68–73.
4. Denke, M.A. (2001) Connections between obesity and dyslipidaemia. *Curr. Opin. Lipidol.*, **12**, 625–628.
5. Rossouw, J.E. (2002) Hormones, genetic factors, and gender differences in cardiovascular disease. *Cardiovasc. Res.*, **53**, 550–557.
6. Betteridge, D.J. (2001) Lipid-lowering trials in diabetes. *Curr. Opin. Lipidol.*, **12**, 619–623.
7. Craig, W.Y., Palomaki, G.E. and Haddow, J.E. (1989) Cigarette smoking and serum lipid and lipoprotein concentrations: an analysis of published data. *Br. Med. J.*, **298**, 784–788.
8. Pearson, T.A. (1996) Alcohol and heart disease. *Circulation*, **94**, 3023–3025.
9. Bernstein, M.S., Costanza, M.C., James, R.W., Morris, M.A., Cambien, F., Raoux, S. and Morabia, A. (2002) Physical activity may modulate effects of ApoE genotype on lipid profile. *Arterioscler. Thromb. Vasc. Biol.*, **22**, 133–140.
10. Couillard, C., Despres, J.P., Lamarche, B., Bergeron, J., Gagnon, J., Leon, A.S., Rao, D.C., Skinner, J.S., Wilmore, J.H. and Bouchard, C. (2001) Effects of endurance exercise training on plasma HDL cholesterol levels depend on levels of triglycerides: evidence from men of the Health, Risk Factors, Exercise Training and Genetics (HERITAGE) Family Study. *Arterioscler. Thromb. Vasc. Biol.*, **21**, 1226–1232.

11. Fielding, C.J. and Fielding, P.E. (1995) Molecular physiology of reverse cholesterol transport. *J. Lipid Res.*, **36**, 211–228.
12. Reichl, D. and Miller, N.E. (1989) Pathophysiology of reverse cholesterol transport. Insights from inherited disorders of lipoprotein metabolism. *Arteriosclerosis*, **9**, 785–797.
13. Rader, D.J. and Maugeais, C. (2000) Genes influencing HDL metabolism: new perspectives and implications for atherosclerosis prevention. *Mol. Med. Today*, **6**, 170–175.
14. Tall, A.R., Jiang, X., Luo, Y. and Silver, D. (2000) 1999 George Lyman Duff memorial lecture: lipid transfer proteins, HDL metabolism, and atherogenesis. *Arterioscler. Thromb. Vasc. Biol.*, **20**, 1185–1188.
15. Glomset, J.A. (1973) The metabolic role of lecithin: cholesterol acyltransferase: perspectives from pathology. *Adv. Lipid Res.*, **11**, 1–65.
16. Clee, S.M., Zwinderman, A.H., Engert, J.C., Zwarts, K.Y., Molhuizen, H.O., Roomp, K., Jukema, J.W., van Wijland, M., van Dam, M., Hudson, T.J. *et al.* (2001) Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. *Circulation*, **103**, 1198–1205.
17. Larson, I.A., Ordovas, J.M., Barnard, J.R., Hoffmann, M.M., Feussner, G., Lamon-Fava, S. and Schaefer, E.J. (2002) Effects of apolipoprotein A–I genetic variations on plasma apolipoprotein, serum lipoprotein and glucose levels. *Clin. Genet.*, **61**, 176–184.
18. Zerba, K.E., Ferrell, R.E. and Sing, C.F. (2000) Complex adaptive systems and human health: the influence of common genotypes of the apolipoprotein E (ApoE) gene polymorphism and age on the relational order within a field of lipid metabolism traits. *Hum. Genet.*, **107**, 466–475.
19. Stengard, J.H., Clark, A.G., Weiss, K.M., Kardia, S., Nickerson, D.A., Salomaa, V., Ehnholm, C., Boerwinkle, E. and Sing, C.F. (2002) Contributions of 18 additional DNA sequence variations in the gene encoding apolipoprotein E to explaining variation in quantitative measures of lipid metabolism. *Am. J. Hum. Genet.*, **71**, 501–517.
20. Corbex, M., Poirier, O., Fumeron, F., Betoulle, D., Evans, A., Ruidavets, J.B., Arveiler, D., Luc, G., Tiret, L. and Cambien, F. (2000) Extensive association analysis between the CETP gene and coronary heart disease phenotypes reveals several putative functional polymorphisms and gene–environment interaction. *Genet. Epidemiol.*, **19**, 64–80.
21. Kuivenhoven, J.A., de Knijff, P., Boer, J.M., Smalheer, H.A., Botma, G.J., Seidell, J.C., Kastelein, J.J. and Pritchard, P.H. (1997) Heterogeneity at the CETP gene locus. Influence on plasma CETP concentrations and HDL cholesterol levels. *Arterioscler. Thromb. Vasc. Biol.*, **17**, 560–568.
22. Gudnason, V., Kakko, S., Nicaud, V., Savolainen, M.J., Kesaniemi, Y.A., Tahvanainen, E. and Humphries, S. (1999) Cholesteryl ester transfer protein gene effect on CETP activity and plasma high-density lipoprotein in European populations. The EARS Group. *Eur. J. Clin. Invest*, **29**, 116–128.
23. deLemos, A.S., Wolfe, M.L., Long, C.J., Sivapackianathan, R. and Rader, D.J. (2002) Identification of genetic variants in endothelial lipase in persons with elevated high-density lipoprotein cholesterol. *Circulation*, **106**, 1321–1326.
24. Cohen, J.C., Vega, G.L. and Grundy, S.M. (1999) Hepatic lipase: new insights from genetic and metabolic studies. *Curr. Opin. Lipidol.*, **10**, 259–267.
25. Miettinen, H.E., Gylling, H., Tenhunen, J., Virtamo, J., Jauhiainen, M., Huttunen, J.K., Kantola, I., Miettinen, T.A. and Kontula, K. (1998) Molecular genetic study of Finns with hypoalphalipoproteinemia and hyperalphalipoproteinemia: a novel Gly230 Arg mutation (LCAT[Fin]) of lecithin:cholesterol acyltransferase (LCAT) accounts for 5% of cases with very low serum HDL cholesterol levels. *Arterioscler. Thromb. Vasc. Biol.*, **18**, 591–598.
26. Wang, J., Freeman, D.J., Grundy, S.M., Levine, D.M., Guerra, R. and Cohen, J.C. (1998) Linkage between cholesterol 7alpha-hydroxylase and high plasma low-density lipoprotein cholesterol concentrations. *J. Clin. Invest*, **101**, 1283–1291.
27. Chen, W., Srinivasan, S.R., Elkasabany, A., Ellsworth, D.L., Boerwinkle, E. and Berenson, G.S. (2001) Influence of lipoprotein lipase serine 447 stop polymorphism on tracking of triglycerides and HDL cholesterol from childhood to adulthood and familial risk of coronary artery disease: the Bogalusa heart study. *Atherosclerosis*, **159**, 367–373.
28. Fisher, R.M., Humphries, S.E. and Talmud, P.J. (1997) Common variation in the lipoprotein lipase gene: effects on plasma lipids and risk of atherosclerosis. *Atherosclerosis*, **135**, 145–159.
29. Bruce, C., Chouinard, R.A. Jr and Tall, A.R. (1998) Plasma lipid transfer proteins, high-density lipoproteins, and reverse cholesterol transport. *A. Rev. Nutr.*, **18**, 297–330.
30. Acton, S., Osgood, D., Donoghue, M., Corella, D., Pocovi, M., Cenarro, A., Mozas, P., Keilty, J., Squazzo, S., Woolf, E.A. and Ordovas, J.M. (1999) Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. *Arterioscler. Thromb. Vasc. Biol.*, **19**, 1734–1743.
31. Tabor, H.K., Risch, N.J. and Myers, R.M. (2002) Opinion: candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet.*, **3**, 391–397.
32. Cardon, L.R. and Bell, J.I. (2001) Association study designs for complex diseases. *Nat. Rev. Genet.*, **2**, 91–99.
33. Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A. and Contopoulos-Ioannidis, D.G. (2001) Replication validity of genetic association studies. *Nat. Genet.*, **29**, 306–309.
34. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. and Hirschhorn, J.N. (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.*, **33**, 177–182.
35. Knoblauch, H., Bauerfeind, A., Krahenbuhl, C., Daury, A., Rohde, K., Bejanin, S., Essioux, L., Schuster, H., Luft, F.C. and Reich, J.G. (2002) Common haplotypes in five genes influence genetic variance of LDL and HDL cholesterol in the general population. *Hum. Mol. Genet.*, **11**, 1477–1485.
36. Hartl D (2000) *A Primer of Population Genetics*. Sinauer Associates, Sunderland, MA.
37. Morabia, A., Cayanis, E., Costanza, M.C., Ross, B.M., Bernstein, M.S., Flaherty, M.S., Alvin, G.B., Das, K., Morris, M.A., Penchaszadeh, G.K., Zhang, P. and Gilliam, T.C. (2003) Association between lipoprotein lipase (LPL) gene and blood lipids: a common variant for a common trait? *Genet. Epidemiol.*, **24**, 309–321.
38. Knoblauch, H., Busjahn, A., Munter,S., Nagy, Z., Faulhaber, H.D., Schuster, H. and Luft, F.C. (1997) Heritability analysis of lipids and three gene loci in twins link the macrophage scavenger receptor to HDL cholesterol concentrations. *Arterioscler. Thromb. Vasc. Biol.*, **17**, 2054–2060.
39. Bernstein, M.S., Costanza, M.C., James, R.W., Morris, M.A., Cambien, F., Raoux, S. and Morabia, A. (2003) No physical activity × CETP 1b.−629 interaction effects on lipid profile. *Med. Sci. Sports Exerc.*, **35**, 1124–1129.
40. Allison, D.B., Heo, M., Schork, N.J., Wong, S.L. and Elston, R.C. (1998) Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power. *Hum. Hered.*, **48**, 97–107.
41. Goring, H.H., Terwilliger, J.D. and Blangero, J. (2001) Large upward bias in estimation of locus-specific effects from genomewide scans. *Am. J. Hum. Genet.*, **69**, 1357–1369.
42. Morabia, A., Bernstein, M., Kumanyika, S., Sorenson, A., Mabiala, I., Prodolliet, B., Rolfo, I. and Luong, B.L. (1994) [Development and validation of a semi-quantitative food questionnaire based on a population survey] [Développement et validation d'un questionnaire alimentaire semi-quantitatif à partir d'une enquête de population.]. *Soz. Praventivmed.*, **39**, 345–369.
43. Bernstein, M., Sloutskis, D., Kumanyika, S., Sparti, A., Schutz, Y. and Morabia, A. (1998) A data-based approach for developing a physical activity frequency questionnaire. *Am. J. Epidemiol.*, **147**, 147–154.
44. Morabia, A., Bernstein, M., Heritier, S. and Ylli, A. (1997) Community-based surveillance of cardiovascular risk factors in Geneva: methods, resulting distributions, and comparisons with other populations. *Prev. Med.*, **26**, 311–319.
45. Friedewald, W.T., Levy, R.I. and Fredrickson, D.S. (1972) Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.*, **18**, 499–502.
46. Kalachikov, S., Evgrafov, O., Ross, B., Winawer, M., Barker-Cummings, C., Boneschi, F.M., Choi, C., Morozov, P., Das, K., Teplitskaya, E. *et al.* (2002) Mutations in LGI1 cause autosomal-dominant partial epilepsy with auditory features. *Nat. Genet.*, **30**, 335–341.
47. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N. *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.*, **22**, 231–238.
48. Kruglyak, L. and Nickerson, D.A. (2001) Variation is the spice of life. *Nat. Genet.*, **27**, 234–236.
49. Bernstein, M.S., Morabia, A. and Sloutskis, D. (1999) Definition and prevalence of sedentarism in an urban population. *Am. J. Public Health*, **89**, 862–867.
50. Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691–692.
51. Anonymous (1999) *S-PLUS 2000 Guide to Statistics*. Insightful Corp., Seattle, WA.